Project Proposal

Solving the EEG Inverse Problem with Deep Learning Methods for Enhanced Brain Activity Localization

Submitted by: Team BrainForge

210173T – GALLAGE D.
210201F – GURUGE S.M.L.
210406M – NANDAKA T.H.T.D.
210574A – SANKALPA L.M.P.

Project Supervisors

Dr. Charith Chitraranjan (Internal)
Dr. Varuna De Silva (External)
Ms. Mindula Illeperuma (External)

Department of Computer Science & Engineering
University of Moratuwa
Sri Lanka

Contents

List of Tables

List of Figures

1 INTRODUCTION

1.1 Background

Electroencephalography (EEG) is a non-invasive method for recording the brain's electrical activity. It does this by measuring voltage changes on the scalp that result from neural activity. EEG is commonly used in neuroscience, for clinical diagnoses like epilepsy and sleep disorders, in cognitive research, and in developing brain-computer interfaces. It offers a high temporal resolution of brain activity and is more affordable and portable than other brain imaging techniques(for more details, see Appendix A).

The EEG Forward Problem involves calculating the expected electrical potentials on the scalp, which is what the EEG sensors detect. This calculation is based on a known distribution of neural electrical sources within the brain and a model of the head's conductive properties [1]. To solve the forward problem, it is important to accurately model the head's shape and conductivity profiles. This helps predict how neural activity translates into scalp potentials. The problem is well-defined and has a unique solution using biophysical equations and numerical methods such as Boundary Element Method (BEM) [2], Finite Element Method (FEM) [3], or Finite Difference Method (FDM) [4]. The forward model creates a lead field matrix that maps source activity to EEG sensor measurements.

The EEG Inverse Problem is the reverse challenge: estimating the location and activity of the neural sources inside the brain from the measured EEG scalp data. It involves estimating the spatial distribution of neuronal current sources inside the brain based on non-invasive EEG signals measured from the scalp. Importantly, this problem is inherently ill-posed, meaning that many different source configurations within the brain can produce identical scalp measurements, making it impossible to determine a unique solution without additional constraints. This ambiguity arises due to the limited number of sensors and the complex conductive properties of head tissues.

Accurate source localization is essential for a wide spectrum of applications, including epilepsy diagnostics, cognitive neuroscience studies, brain-computer interfaces (BCIs), and neurofeedback systems. Consequently, a variety of traditional source localization methods have been developed to tackle the inverse problem. Classical algorithms like Weighted Minimum Norm Estimation (wMNE) [5], Exact Low Resolution Brain Electromagnetic Tomography (eLORETA) [6], and the Linearly Constrained Minimum Variance (LCMV) [6] Beamformer introduce anatomical and biophysical priors to regularize the solution space and yield physiologically plausible estimates.

Despite their practicality and computational efficiency, these traditional approaches have notable limitations:

- Low spatial resolution: These methods tend to blur or merge nearby sources, which can obscure subtle yet clinically significant patterns of neural activity, ultimately impacting diagnostic precision and the ability to localize brain functions accurately for research.

- Dependence on noise characteristics: Their performance is often sensitive to sensor noise, artifacts, and signal-to-noise ratio (SNR), which can vary widely across EEG recordings.
- Sensitivity to model parameters: Estimates can be strongly influenced by choices such as regularization strength, the accuracy of head models, or electrode placement, potentially limiting reliability in real-world settings.

To overcome these challenges, recent research has explored deep learning approaches for EEG source localization. Unlike traditional models, deep learning methods can learn complex, non-linear mappings between scalp potentials and underlying sources directly from large-scale data, potentially capturing intricacies that hand-crafted algorithms miss. For instance:

- ConvDip [7] employs convolutional neural networks (CNNs) to automatically extract spatial features from EEG data, allowing for data-driven inference of underlying brain sources.
- DeepSIF [8] extends this idea by combining convolutional and recurrent neural networks to simultaneously capture both spatial and temporal dynamics present in EEG recordings, offering a more holistic analysis of brain activity.

Among these, DeepSIF has demonstrated good accuracy and generalization performance, outperforming both conventional approaches and earlier deep learning models across multiple benchmark datasets. This trajectory of research highlights the promise of data-driven frameworks in addressing the longstanding challenges of EEG source localization.

1.2 Motivation

Current advanced deep learning methods for EEG source localization still face several significant limitations and challenges that restrict their broader clinical and scientific impact.

First, many existing approaches focus primarily on the outer cortical layer and do not adequately account for subcortical brain regions. However, deep brain structures—such as the thalamus and hippocampus—can play a crucial role in various neurological disorders. Omitting these regions may lead to incomplete or less accurate source localization, limiting the utility of these methods in conditions involving subcortical activity.

Second, the generation of synthetic data for training deep learning models is extremely resource-intensive. Current datasets can require on the order of terabytes of storage space. This translates into high computational costs and substantial storage requirements, which may render these methods inaccessible to research groups or clinicians with limited computing infrastructure. Such barriers can slow down research progress and hinder widespread adoption in resource-constrained environments.

Third, these techniques often focus on averaged interictal spikes in EEG recordings. While averaging reduces noise and inter-trial variability, it risks obscuring important information present in individual spikes. In clinical practice, the morphology, timing, and spatial origin of single interictal spikes are critical for accurate diagnosis and treatment planning in epilepsy. Averaging can mask transient or atypical events, potentially limiting the robustness and clinical relevance of models when applied to diverse patient data.

Fourth motivation for this research is the exploration of new deep learning architectures for EEG source localization. Existing methods have not yet combined CNNs with Transformer-based models, leaving an opportunity to enhance performance. CNNs are well-suited for capturing spatial features from EEG data [9], offering an advantage over traditional fully connected architectures. Meanwhile, Transformers [10] demonstrate superior accuracy when trained on large datasets [11]. By integrating these two approaches, the proposed architecture is expected to improve accuracy and robustness compared to previous methods.

Moreover, many spike imaging approaches rely on identifying peak activations and segmenting active brain regions using threshold-based methods. While effective to some extent, this process often ignores the temporal progression and dynamics across the full EEG time series. Fully leveraging temporal information could substantially improve the accuracy and temporal coherence of source localization, especially for transient or non-stationary brain activity.

Addressing these challenges promises to deliver more clinically relevant and scientifically robust EEG source localization solutions. Incorporating subcortical regions will provide a more comprehensive understanding of brain activity, beneficial for disorders involving deep brain structures. Reducing data generation and storage demands will enhance accessibility for a broader research and clinical community. Finally, analyzing individual spike events and utilizing the full temporal EEG data are expected to improve model sensitivity and specificity, helping translate computational advances into meaningful clinical impact.

1.3 Problem Statement

The EEG inverse problem pinpointing the exact neural sources from scalp-recorded brain signals remains one of the most challenging tasks in neuroimaging. Traditional classical methods, such as equivalent current dipole fitting or minimum-norm estimates, have been widely used for decades. However, they often struggle with accuracy and spatial resolution due to their reliance on idealized head models and assumptions that rarely hold in real-world data [5] [6]. These methods can be sensitive to modeling errors, such as inaccuracies in tissue conductivities or electrode positions, and they tend to produce biased solutions that favor superficial cortical sources over deeper ones. Furthermore, they can be computationally expensive when working with high-density EEG setups and large source spaces [12], and they often lack robust mechanisms for quantifying uncertainty making their results harder to interpret and trust in clinical or research settings [13].

Deep learning-based approaches have emerged as a promising alternative, offering the potential for rapid, data-driven source localization without the need for iterative optimization. However, these models introduce their own set of limitations. They typically require vast amounts of training data [14], which, in the absence of large annotated EEG datasets, is often generated synthetically through forward simulations. This synthetic data generation process is highly resource-intensive [15] and may not fully capture the complexity and variability of real-world EEG signals. As a result, there is often a domain gap between synthetic and real data [16] , leading to poor generalizability when models are applied to new subjects, electrode configurations, or noise conditions not represented in the training set.

Another drawback of current deep learning methods is their tendency to overfit to specific experimental setups or priors used during simulation [17]. This makes them vulnerable to performance degradation in real-world scenarios, especially when the data deviates from the conditions under which the models were trained. Moreover, many of these approaches function as "black boxes," providing little insight into their decision-making processes. This lack of interpretability and uncertainty estimation is a significant concern for high-stakes applications such as epilepsy diagnosis, surgical planning, and brain–computer interfaces, where understanding the reliability of localization results is as important as the results themselves.

To tackle these challenges, we suggest a new approach that emphasizes synthetic data generation and deep learning model design. Our method seeks to lower the resource demands of current data generation processes while creating more realistic datasets by using multiple source activations at the same time. We also intend to create a new deep learning architecture aimed at improving localization accuracy and reducing overfitting. This will make the model more reliable and effective across different subjects, electrode setups, and noise conditions. Through this research, we aim to develop a solution that not only achieves high accuracy and robustness but also offers improved generalizability across subjects and experimental conditions. This architecture will be evaluated on both simulated and real EEG datasets, with a focus on its ability to handle noisy, heterogeneous data while maintaining consistent performance. Ultimately, our goal is to provide a reliable, interpretable, and computationally efficient tool for EEG source localization that advances both clinical and research applications.

1.4 Research Objectives

In this work, we propose several enhancements to address these limitations: Deep brain region synthetic data generation: Employing neural mass models to simulate activity in deeper brain regions.

1. **Address inefficiencies in synthetic data generation:**
   A major challenge in deep learning-based EEG source localization is the reliance on large-scale synthetic datasets generated through forward simulations. Traditional

approaches to data generation are highly resource-intensive, requiring nearly 6 TB of storage and about 4000 hours of computation to prepare training datasets [15, 14]. Such requirements significantly limit scalability and practical applicability. Our objective is to overcome these inefficiencies by introducing a more optimized pipeline that reduces the storage requirement to approximately 30 GB and computation time to 900 hours, thereby making synthetic data generation considerably more practical and accessible for large-scale research.

2. **Propose a novel deep learning architecture:**
   While convolutional neural networks (CNNs) have been successfully applied for extracting spatial features from EEG data [18], Transformers have shown strong performance on large-scale datasets compared to other neural architectures [11]. However, no prior work has explored a hybrid combination of these two architectures for the EEG inverse problem. To address this gap, we propose a hybrid CNN–Transformer model that couples a CNN-based spatial module for robust feature extraction with a Transformer-based temporal module designed to capture complex temporal dependencies. This novel architecture is expected to provide a more powerful representation of EEG signals, enabling improved localization performance.

3. **Improve accuracy and efficiency in EEG source localization:**
   Beyond efficiency in data generation and novelty in architecture design, a core objective of this research is to enhance the accuracy and robustness of EEG source localization. Traditional optimization-based methods often fail under noisy conditions and are computationally demanding [5]. Recent deep learning approaches have improved performance but still suffer from scalability and generalization issues [19, 16]. Our proposed hybrid approach aims to deliver higher accuracy even in noisy, real-world settings, while simultaneously reducing computational and storage overhead. This balance of accuracy and efficiency is crucial for making EEG source localization methods more practical for clinical and research applications.

4. **Incorporate multiple source activations in source imaging**
   In real-world situations, EEG signals usually come from the simultaneous activation of several neural sources instead of one focal generator. This is especially true in cases like epilepsy or during complex cognitive tasks [20]. To improve the ecological validity and reliability of source localization, our research aims to model and evaluate multiple-source activations in the proposed deep learning framework.

These proposed modifications aim to extend applicability, improve accuracy, and reduce the resource overhead required for model training.

2 LITERATURE REVIEW

Recent studies have looked into different computational methods to improve the accuracy and reliability of source localization in EEG analysis. Progress in mathematical modeling, signal processing, and machine learning has led to better reconstruction of brain activity patterns. Techniques range from traditional inverse solution algorithms to modern deep learning frameworks. The literature points out the trade-offs among spatial resolution, resistance to noise, and computational efficiency. It also considers factors like electrode density, head model accuracy, and connectivity measures. These developments offer a solid basis for choosing and adjusting appropriate methods for analysis of brain activity patterns using EEG.

2.1 Traditional Mathematical Model Based Solutions

Traditional mathematical models for solving the EEG inverse problem depend on physical and mathematical representations of the head and brain activity. These methods generally use a forward model to show how neural currents create scalp potentials. They then apply inverse techniques to estimate where the sources are located and their strengths. These models include assumptions about source distributions, head shape, and measures to handle the uncertain nature of the problem. Two common assumptions are that sources project perpendicularly to the cortex surface, which keeps dipole orientations aligned with the cortical shape, and that sources come from the center of a specific area, which simplifies the spatial representation of neural activity [6].

Common approaches include minimum norm estimates, dipole modeling, and spatial filtering techniques. These methods are efficient and easy to understand, but they often rely on simplifying assumptions. These assumptions can limit their accuracy in complex situations that involve distributed or non-orthogonal source configurations.

The Table 1 summarizes key mathematical-based architectures, including Minimum Norm Estimation (MNE), WMNE, eLORETA, Standardized Low-Resolution Electromagnetic Tomography (sLORETA), and Dynamic Statistical Parametric Mapping (dSPM).

Table 1: Comparison between traditional mathematical models

Model	**Focus**	**Strengths**	**Weaknesses**	**Best Used For**
MNE (Minimum Norm Estimation) [5]	Estimates source activity by minimizing the L2-norm of the source amplitudes, assuming distributed sources.	Simple, computationally efficient, provides smooth solutions across the cortex.	Tends to favor superficial sources, sensitive to noise, may produce overly diffuse solutions.	General-purpose source localization when computational speed is prioritized over high spatial resolution.
WMNE (Weighted Minimum Norm Estimation) [5]	Extends MNE by applying depth-weighting to compensate for bias toward superficial sources.	Reduces bias toward superficial sources, improves localization of deeper sources.	Still sensitive to noise, requires careful tuning of weighting parameters.	Scenarios requiring improved depth localization with distributed source models.
eLORETA (exact Low-Resolution Electromagnetic Tomography) [6]	Minimizes a cost function with spatial smoothness constraints, ensuring zero localization error for point sources.	High localization accuracy for point sources, robust to noise, enforces spatial coherence.	Low spatial resolution, computationally intensive, assumes linear source models.	Applications requiring precise localization of focal sources, such as epilepsy studies.
sLORETA (Standardized Low-Resolution Electromagnetic Tomography) [21]	Standardizes source estimates to reduce amplitude bias, focusing on localization accuracy.	Improved localization over MNE, robust to noise, computationally efficient.	Limited resolution, may struggle with complex or distributed sources.	Functional connectivity studies and localization of focal brain activity.
dSPM (Dynamic Statistical Parametric Mapping) [22]	Combines MNE with statistical normalization to map source activity relative to noise levels.	Enhances interpretability by providing statistical maps, robust to noise variations.	Dependent on accurate noise covariance estimation, may overestimate source extent.	Neuroimaging studies requiring statistical interpretation of source activity.

2.2 EEG source localization analysis steps using the academic application Cartool

Cartool [23] is an academic software tool that many use for EEG source localization. It is known for its complete processing pipeline, which covers preprocessing, head model construction, and inverse solution computation. It integrates three linear distributed source models MNE, LORETA, and LAURA (Local Autoregressive Average) which offer complementary approaches to solving the EEG inverse problem with varying spatial constraints and regularization strategies.

2.2.1 Preprocessing steps for EEG data

Robust preprocessing of EEG data is fundamental for reliable source localization. Cartool implements a sequence of preprocessing steps:

- Temporal Filtering: Employs non-causal, 2nd order IIR Butterworth filters to remove non-physiological frequency components while preserving neural oscillations vital for source analysis.
- Down Sampling: Reduces data volume and computational load by focusing processing on relevant frequency bands.
- Artifact Correction: Utilizes Independent Component Analysis (ICA) to identify and mitigate artifacts such as eye blinks, muscle activity, and cardiac signals, thereby minimizing contamination of neural signals.
- Spatial Filtering: Detects and smooths transient electrode outliers through spatial filtering, preserving the topographic integrity of the EEG maps.
- Epoch Selection: Uses statistical measures (variance, skewness, absolute amplitude) to flag and exclude contaminated epochs, ensuring clean data inputs for source modeling, a step shown to enhance model precision.

2.2.2 Main steps of source localization

The principal workflow for source localization after the data preprocessing within Cartool involves:

- Head Model Construction: Cartool uses individual MRI data to create accurate head models. These models include the skull, scalp, and brain tissues. Key processing steps involve removing the skull, correcting bias fields, and dividing gray matter. These steps improve the accuracy compared to generic or spherical head models.
- Solution Point Definition: Thousands of solution points are spread evenly across the segmented gray matter volume. This improves spatial coverage and lowers localization bias.

- Lead Field Calculation: The software calculates lead fields using methods like the Locally Spherical Model with Anatomical Constraints (LSMAC), Boundary Element Method (BEM), and Finite Element Method (FEM). This models how cortical sources project to scalp potentials while maintaining anatomical realism.
- Inverse Solution Computation: Three linear distributed source models are implemented:
  1. MNE: Provides smoothed source distributions minimizing overall current amplitude, though it may underestimate deep sources.
  2. LORETA: Promotes spatial smoothness to enhance robustness at the expense of spatial resolution.
  3. LAURA: Incorporates adaptive spatial weighting for improved source discrimination.

Normalization and regularization techniques are applied to balance noise reduction and spatial resolution, a critical but challenging aspect of inverse modeling.

2.3 Deep Learning-Based Architectures

2.3.1 ConvDip

Among the first and most influential deep learning models for EEG source imaging is ConvDip, a Convolutional Neural Network (CNN) proposed by Hecker et al. (2021) [7]. The model was a pioneering achievement, as it was one of the first to solve the inverse problem for a distributed dipole model with multiple sources, a task where previous neural networks had been limited to only one or two sources. By demonstrating that a CNN could generate comprehensive source images from a single snapshot of EEG data, ConvDip established a new benchmark for data-driven source localization.

Architecture and Methodology

The high-level architecture of ConvDip is shown in Fig. 1. This "moderately shallow" CNN was designed specifically for the source localization task. The network takes a single time-point of EEG data, interpolated into a 2D image, as its input. This input is processed through a single convolutional layer, a fully-connected layer, and an output layer with 5,124 neurons, where each neuron corresponds to a voxel in the brain's source model. A key design choice was the deliberate omission of pooling layers, which are common in image classification but would discard the precise spatial information that is essential for a "where-task" like source localization.

The model's training is entirely data-driven, relying on 100,000 simulated EEG samples. These samples were generated to mimic smooth, circular source patches, and realistic noise from actual EEG was added to improve robustness. Instead of a standard Mean Squared Error, ConvDip was trained using the Weighted Hausdorff Distance (WHD) as a loss function. This metric prioritizes minimizing the positional error between the predicted and

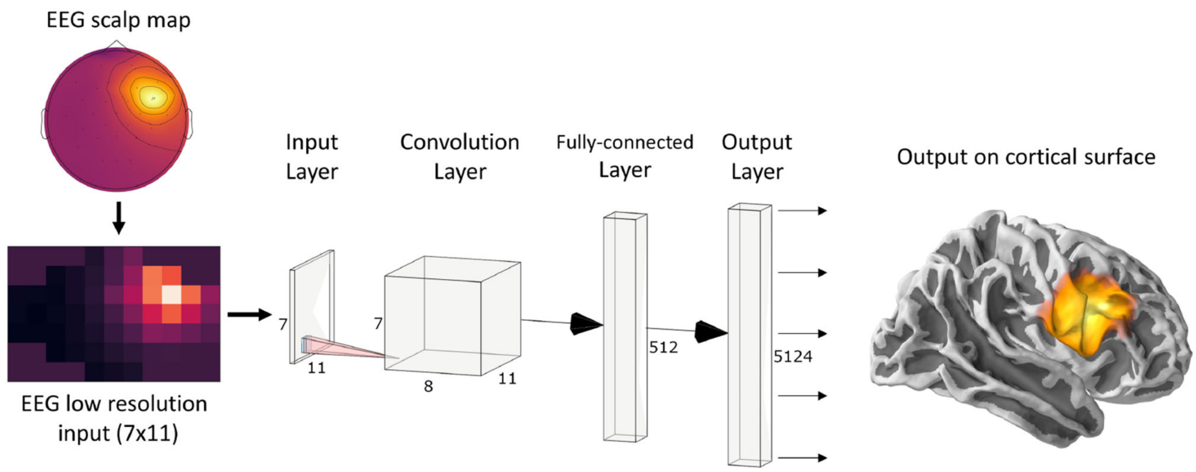

Fig. 1: High-Level architecture diagram of ConvDip

true source locations, effectively separating the challenge of localization from the estimation of source amplitude. A post-hoc scaling step is used to determine the final source intensity.

Performance and Key Advantages

When evaluated against classical methods like eLORETA, LCMV, and cMEM, ConvDip demonstrated superior performance across most metrics in both single and multiple-source simulations. It achieved a higher Area Under the ROC Curve (AUC) and lower normalized Mean Squared Error (nMSE), indicating a better ability to estimate the source's spatial extent and global similarity.

ConvDip possesses several key advantages over traditional approaches:

- Computational Speed: Once trained, the model is remarkably fast, computing an inverse solution in under 40 ms on a GPU. This is orders of magnitude faster than iterative methods like cMEM ($\approx$ 11.5 s) and significantly faster than eLORETA ($\approx$ 129 ms) and LCMV ($\approx$ 98 ms), making it suitable for real-time applications like neurofeedback and BCIs.

- Deep Source Localization: Classical methods are often biased towards superficial sources, struggling to localize activity deep within the brain. ConvDip shows almost no drop-off in localization accuracy with increasing source depth, learning to recognize the scalp patterns of deep sources from its training data.

- Robustness: The model was rigorously tested by training it on data from one forward model and evaluating it on data from another, a method that avoids the "inverse crime." Its strong performance in this scenario proves its robustness against the inevitable inaccuracies in head models used in real-world applications.

Limitations and Future Directions

Despite its groundbreaking performance, ConvDip has several limitations that highlight directions for future research.

- Temporal Blindness: The model works with a single, fixed time point of EEG. It completely ignores the complex changes in the neural signal over time. This means it cannot use data from a baseline period to estimate noise levels or take advantage of how the signal changes as time progresses.
- Training Data Bias: The model's performance is fundamentally tied to its training data. Its reliance on simple, circular source shapes raises questions about its ability to generalize to sources with more complex or elongated geometries.
- Training Cost: While prediction is fast, training the model is computationally expensive, requiring over 10 hours on a high-end GPU for 100,000 samples. This poses a practical barrier to creating subject-specific models based on individual anatomies.

These limitations offer a strong reason to develop better architectures. The model's difficulty with time specifically encourages the development of spatiotemporal models that can handle EEG time series. At the same time, the issues with training data dependency and cost push research toward more efficient and reliable solutions.

2.3.2 DeepSIF

The Deep learning-based Source Imaging Framework (DeepSIF) represents a significant evolution from static, single-timepoint models like ConvDip, introducing a dynamic, spatiotemporal approach to Electrophysiological Source Imaging (ESI) [7, 8]. Its performance is driven not only by its advanced network architecture but also by a comprehensive synthetic data generation pipeline designed to overcome the scarcity of ground-truth physiological data. By simulating realistic brain dynamics, DeepSIF addresses key limitations of both classical methods and early deep learning approaches.

Synthetic Data Generation

To overcome the scarcity of ground-truth data in physiological studies, DeepSIF employs a robust three-stage pipeline to generate realistic synthetic EEG datasets for training and testing. This process relies on neural mass models (NMM) to simulate plausible spatiotemporal brain dynamics.

- Stage 1: Raw Data Simulation The first stage utilizes The Virtual Brain (TVB) library and the Jansen-Rit model to generate raw time-series data. The simulation is configured using a connectivity matrix of 998 brain regions, where the Jansen-Rit model produces neural activity representing the difference between excitatory and inhibitory potentials. To ensure diversity, region-specific parameters (such as coupling strength) are varied, and simulations are executed in 200-second chunks per region with a step size of 0.5ms.

- Stage 2: Spike Extraction and Preprocessing In the second stage, the raw simulated data undergoes preprocessing to isolate high-quality spike waveforms. The data is downsampled by a factor of four to reduce computational load, and initial transient phases (the first 1000 samples) are discarded to ensure stability. A rigorous selection logic is then applied: candidate spikes must meet an amplitude threshold of 8 and satisfy specific isolation rules. These rules ensure that the spike originates strictly from the target region and is temporally isolated, requiring a 900-sample window free of significant activity from other brain regions.

- Stage 3: Synthetic Source Construction The final stage involves constructing the training dataset by generating synthetic source patches. Using the fsaverage5 atlas, brain regions are grouped into patches to represent spatially contiguous neural activity. The previously extracted NMM spikes are mapped to these patches and scaled to achieve varying Signal-to-Noise Ratios (SNRs) between 5 and 20 dB. To simulate realistic EEG recordings, the source activity is projected to sensor space (75 electrodes) using a leadfield matrix, and Gaussian noise is added before normalization.

Architecture and Training

DeepSIF's network architecture is composed of two distinct modules designed to tackle the ESI problem sequentially, as illustrated in Fig. 2:

- Spatial Module: This component uses a residual network (ResNet) with fully connected layers to process spatial information from scalp EEG data at each time step. Its main function is to learn the complex, non-linear relationship between the smeared scalp topography and an initial estimate of the source distribution. This process starts to reverse the distortion caused by volume conduction.

- Temporal Module: The output from the spatial module is then fed into a temporal module, which employs Long Short-Term Memory (LSTM) recurrent layers. This module aggregates the spatially-processed information over time, learning the inherent temporal patterns and dependencies in the brain signal to reconstruct a smooth and dynamically plausible "movie" of brain activity.

This spatiotemporal design is a major advancement, as it allows the model to image the full evolution of brain dynamics, unlike static models that produce only a single snapshot.

A key part of DeepSIF is its use of very realistic synthetic training data. Unlike the basic patch-source models from earlier methods, DeepSIF uses interconnected Neural Mass Models (NMMs) in The Virtual Brain simulator. These NMMs can create a wide range of biophysically plausible brain signals, such as resting-state oscillations and pathological events like epileptiform spikes. This gives the network more realistic examples to learn from.

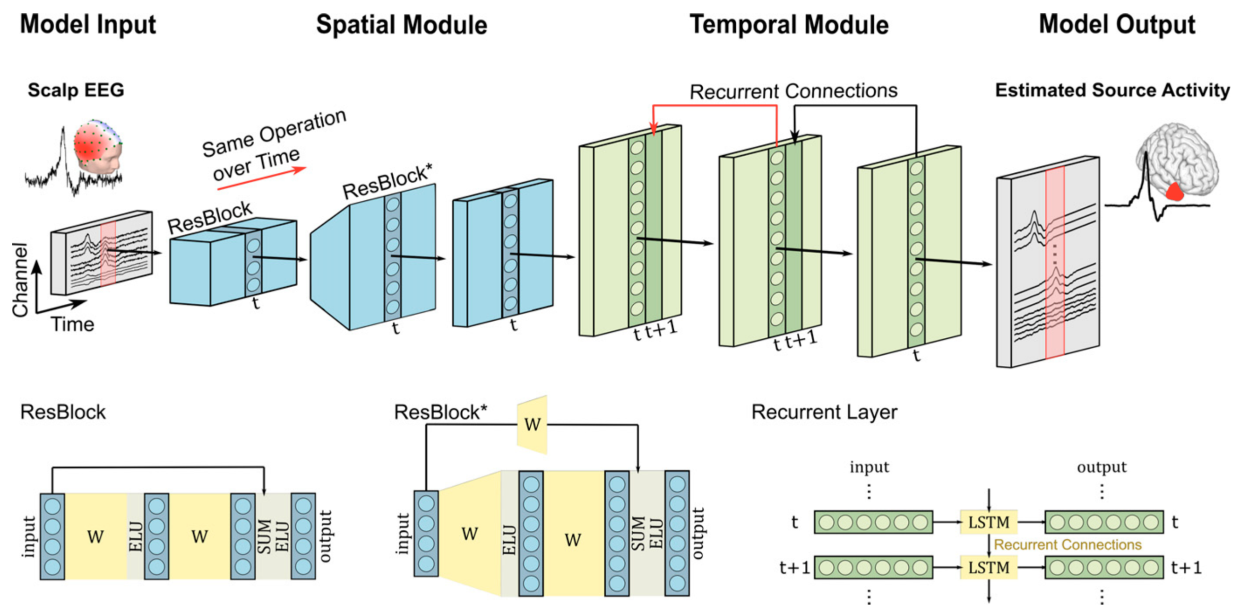

Fig. 2: High-Level architecture diagram of DeepSIF

Key Enhancements and Applications

Following its initial development, the DeepSIF framework has been extended and refined to address critical real-world challenges, enhancing its clinical viability.

- Robustness with Low-Density EEG: A major barrier to the clinical adoption of ESI is the perception that it requires high-density EEG for accurate results, a significant hurdle since standard clinical practice often uses low-density setups (e.g., 19-32 channels). A 2025 study by Rong et al. directly addressed this by investigating the robustness of DeepSIF across varying electrode configurations [24]. The study compared DeepSIF against conventional methods (sLORETA, LCMV) using channel counts ranging from 75 down to 16 in a cohort of 27 drug-resistant epilepsy patients. The results showed that while the performance of conventional methods degraded substantially with fewer electrodes, DeepSIF maintained a high degree of accuracy. With 75 channels, DeepSIF's average spatial dispersion was 7.9 mm; with only 16 channels, this error increased only slightly to 9.0 mm. In stark contrast, sLORETA's error increased from 21.9 mm to 28.1 mm, and LCMV's from 20.0 mm to 28.9 mm under the same conditions. This demonstrates DeepSIF's utility in standard clinical environments using common EEG hardware, lowering the barrier for its adoption as a routine diagnostic tool.

- Personalized DeepSIF (PDeepSIF) for MEG: To address anatomical variability between individuals, the Personalized DeepSIF (PDeepSIF) model was developed for magnetoencephalography (MEG) data [25]. This approach uses a fine-tuning procedure where a generic, pre-trained DeepSIF model is adapted using a small set of synthetic data generated from the patient's own MRI. This personalization significantly improves accuracy, more than halving the spatial dispersion error (from 21.90 mm to 8.19 mm) in a cohort of 29 epilepsy patients and increasing sublobar concordance with clinical findings from 83% to 93%.

- Imaging Ictal (Seizure) Dynamics: The framework was successfully adapted to image ongoing seizure activity, known as ictal dynamics. By training the model on NMMs specifically configured to simulate the rhythmic, evolving patterns of seizures, the specialized DeepSIF model was able to accurately localize ictal sources in a cohort of 33 epilepsy patients, achieving a high spatial specificity of 96% and a low spatial dispersion of 3.80 mm compared to the surgical resection area [26].

Limitations and Future Directions

Despite its significant advancements, the performance of DeepSIF is fundamentally limited by the quality of its synthetic training data. This issue is often called the "sim-to-real" gap. Any difference between the NMM-generated signals and the complex nature of real neural activity can result in poor performance. The original authors pointed out that the model might have overlooked a weak frontal source in a P300 experiment. This may have happened because the training data did not fully represent that specific type of physiological activity. Likewise, major differences in patients' anatomy, such as those due to previous surgeries or large lesions, may need the model to be retrained or fine-tuned to ensure accuracy.

These limitations have illuminated a clear path forward for the field, catalyzing research into new methodologies that aim to close the sim-to-real gap and build upon DeepSIF's foundation.

- Clinical Translation with Low-Density EEG: The demonstrated robustness of DeepSIF with low-density electrode configurations, as shown in the work by Rong et al. [24], is a critical step toward widespread clinical adoption. A key future direction is the development and validation of standardized clinical ESI protocols that leverage this capability. This would allow for the integration of advanced source imaging into routine clinical workflows without requiring specialized high-density hardware, making the technology more accessible and cost-effective.

- Hybrid Physics-Informed Models: A promising direction is the development of hybrid models that integrate the strengths of both classical and deep learning techniques. As proposed by Wegener et al. in their 3D-PIUNet model [27], these methods first use a traditional, physics-based algorithm (e.g., eLORETA) to generate an initial, physically plausible source estimate. A deep neural network is then trained only to refine this estimate, learning to correct its errors based on data-driven priors. This "best-of-both-worlds" approach reduces the dependency on the synthetic simulator by grounding the solution in a physics-informed starting point.

- EEG Foundation Models: Inspired by successes in natural language processing, researchers are developing large-scale "foundation models" for EEG, such as Neuro-GPT by Cui et al. [28] and LaBraM by Jiang et al. [29]. These models are pre-trained on massive and diverse datasets of real EEG recordings, with the goal of learning universal representations of brain activity. These powerful, pre-trained models can then be fine-tuned for specific downstream tasks like source imaging, aiming to minimize the reliance on synthetic data altogether.

- Generative AI for Data Super-Resolution: Another innovative approach uses generative AI to enhance data quality at the sensor level. As demonstrated by Wang et al. [30], diffusion models like STAD (Spatio-Temporal Adaptive Diffusion) are being developed for EEG "super-resolution", learning to generate high-density EEG data (e.g., 256 channels) from standard low-density recordings. This synthetically enhanced data could then be fed into a source imaging algorithm like DeepSIF, potentially boosting its performance and overcoming hardware limitations in data acquisition.

In conclusion, the DeepSIF framework marked a pivotal step in ESI by demonstrating the power of a spatiotemporal, data-driven approach fueled by biophysically realistic simulations. Its successes and inherent limitations have, in turn, become a catalyst for the next generation of research, pushing the field toward hybrid models and large-scale foundation models that promise even greater accuracy and clinical utility.

2.3.3 GRU-based Temporal Modeling

Building upon the foundations established by ConvDip [7] and DeepSIF [8], researchers have explored alternative recurrent architectures for temporal modeling in EEG source localization. A notable contribution in this area is the work by Khosravi et al. [31], who proposed a novel GRU-based neural network architecture that addresses temporal dependencies while maintaining computational efficiency and noise robustness.

Architecture and Methodology

The GRU-based approach presented by Khosravi et al. introduces a dual-path neural network architecture that combines the computational efficiency of Gated Recurrent Units (GRU) with an encoder-decoder framework for enhanced denoising capabilities [31]. As shown in Fig. 3, the model architecture consists of two complementary pathways inspired by autoencoder designs proposed by Cho et al. [32]:

- Path 1 (Direct Mapping): A straightforward pathway comprising two fully connected layers that directly maps electrode measurements to source estimates. The input layer matches the number of electrodes (64 channels), while the output layer corresponds to the number of sources (1260 sources).

- Path 2 (Temporal Processing): A more sophisticated pathway incorporating a fully connected layer followed by two bidirectional GRU networks, each containing 32 forward and 32 backward cells. This encoder-decoder structure learns latent-space representations to improve denoising and extract temporal dynamics from EEG signals.

The final output is produced by element-wise multiplication of the outputs from both paths, allowing the temporal pathway to refine the estimates from the direct mapping pathway.

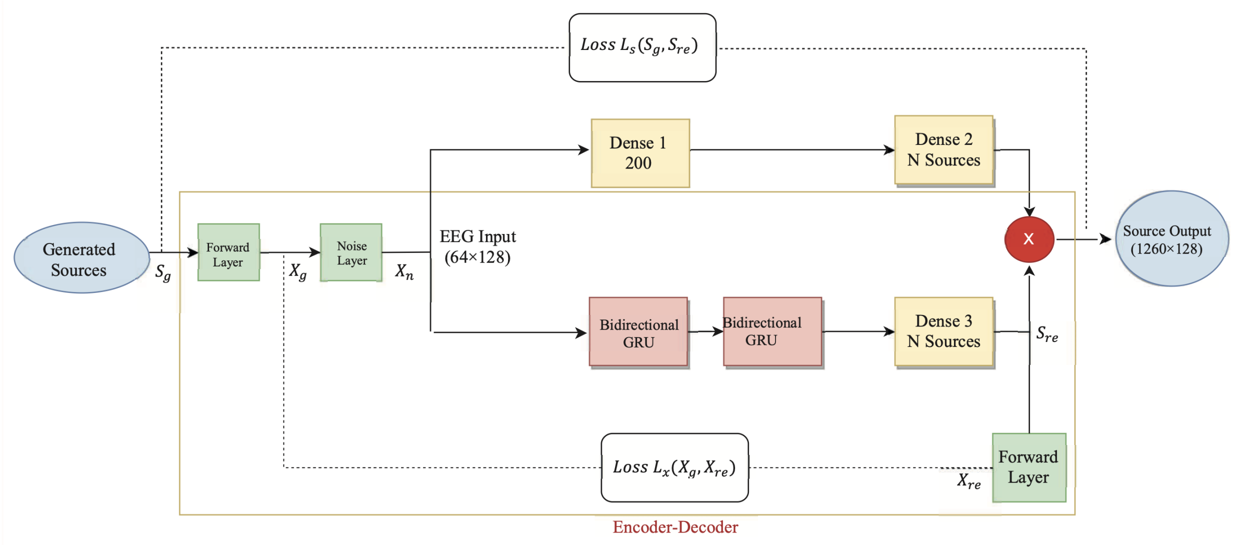

Fig. 3: High-Level architecture diagram of GRU-based encoder-decoder framework

The network employs tanh activation functions for hidden layers and sigmoid for output layers, with dropout (0.2) applied after each hidden layer to prevent overfitting.

Data Generation and Training

The model was trained on 12,000 simulated data samples with a 90%-10% train-validation split. The training data generation process involved several key innovations based on established EEG simulation principles:

- Realistic Source Simulation: Sources were modeled as regions from 116 anatomically defined areas (using automated anatomical labeling), with 1–10 randomly selected active regions per sample.
- Frequency Spectrum Modeling: Source signals followed a $1/f^{\beta}$ frequency spectrum ($0.5 < \beta < 1.5$) to match the natural characteristics of EEG signals.
- Noise Modeling: Realistic noise with varying signal-to-noise ratios (2–10 dB) was added to simulate real-world recording conditions.

The cost function uniquely combines cosine similarity between actual and predicted source vectors with RMS error between clean and reconstructed EEG signals, providing a comprehensive optimization target that addresses both spatial patterns and signal reconstruction quality.

Performance Evaluation and Results

The GRU-based model demonstrated superior performance compared to classical methods (eLORETA and LCMV beamformer) across multiple evaluation metrics, as shown in Tables 2 and 3:

Advantages and Limitations

Table 2: Single Source Localization Performance

Metric	GRU	eLORETA	LCMV
Normalized Mean Square Error (nMSE)	0.00094	0.0184	0.0106
Area Under Curve (AUC)	99.92%	83.63%	63.58%
Mean Localization Error (MLE)	4.87 mm	10.26 mm	18.70 mm

Table 3: Multiple Source Localization Performance

Metric	GRU	eLORETA	LCMV
Normalized Mean Square Error (nMSE)	0.00498	0.0831	0.0462
Area Under Curve (AUC)	85.96%	70.46%	62.87%
Mean Localization Error (MLE)	8.76 mm	20.04 mm	22.86 mm

The GRU-based approach demonstrates several key advantages that distinguish it from alternative methods. In terms of computational efficiency, GRU networks require fewer parameters than LSTM networks while maintaining temporal modeling capabilities [32], making them more practical for deployment in resource constrained environments. The encoder-decoder framework with dual-path architecture provides enhanced noise robustness capabilities, allowing the model to perform reliably under challenging signal conditions commonly encountered in clinical settings. Furthermore, the bidirectional GRU networks effectively capture both past and future temporal relationships, enabling more comprehensive temporal dependency modeling compared to unidirectional approaches. The clinical validation studies demonstrated that the model produces physiologically realistic results, with activation patterns that align consistently with expected neuroanatomy.

Despite these advantages, the GRU-based approach faces several notable limitations that constrain its practical applications. The bidirectional processing architecture, while beneficial for accuracy, prevents real-time applications since future time points must be available before processing can begin. This limitation significantly restricts its utility in brain-computer interface applications and real-time neurofeedback systems. Additionally, the model's performance remains fundamentally limited by the quality and diversity of simulated training data, creating challenges in generalizing to novel patient populations or recording conditions not represented in the training set. The dual-path architecture, while improving accuracy, increases computational complexity during inference, potentially limiting its deployment on mobile or embedded systems where computational resources are constrained.

2.4 CNN-Based Methods for Spatial Feature Extraction

The application of Convolutional Neural Networks (CNNs) to analyze EEG data marks a significant paradigm shift from traditional signal processing to data-driven feature learning. A primary challenge in using CNNs, which are designed for grid-like data such as images, is that EEG recordings come from a sparse and irregular arrangement of scalp electrodes. The common and effective solution is to convert these sparse measurements into a dense, 2D topographical map through projection and interpolation. This creates an image-like representation for each time point, enabling the powerful pattern-recognition capabilities of CNNs to be leveraged for learning complex spatial features from the scalp's electrical field.

The foundational viability of using CNNs for EEG was robustly demonstrated in the work by Schirrmeister et al. [18]. This study showed that deep CNNs could be used effectively for EEG decoding and visualization. The critical insight from their research was that these networks are capable of learning meaningful and discriminative features directly from raw EEG signals, thereby circumventing the need for extensive, manual feature engineering that characterized earlier machine learning approaches. By proving that CNNs could learn relevant spatio-temporal features automatically, this work established the feasibility of applying architectures originally designed for computer vision to the complex domain of electrophysiology, paving the way for their use in more advanced tasks.

Building on this foundation, subsequent research has focused on optimizing the input representations for these networks. The work by Liang et al. [9], for instance, specifically addresses the creation of a "topographic representation module" for use with CNNs in the context of brain-computer interfaces. This line of research is directly relevant as it validates the approach of treating EEG data as a sequence of topographical images and explores how to best structure these maps to enhance the spatial feature extraction performed by the convolutional layers. This confirms that the initial 2D conversion is not only a preprocessing step but also a critical component of the model design that directly impacts performance.

These principles culminated in the development of ConvDip [7], a pioneering CNN architecture designed specifically to solve the EEG inverse problem. The ConvDip model demonstrated that a relatively shallow CNN could learn to map a single, 2D-interpolated snapshot of EEG scalp activity to a distributed dipole source model in the brain. A key architectural innovation highlighted in the study was the deliberate omission of pooling layers. While pooling is standard in image classification to create invariance to object position, it was recognized as detrimental for source localization, where preserving precise location information is essential. This adaptation of the standard CNN blueprint for the specific neuro-imaging problem underscores a critical design principle.

2.5 Source Imaging of Deep-Brain Activities

Standard EEG source imaging methods often struggle to detect and localize deep brain activity, such as that from the thalamus or hippocampus because these signals are inherently weak and are easily masked by stronger cortical activity and background noise.

The Regional Spatiotemporal Kalman Filter (RSTKF) [33] advances beyond conventional spatiotemporal Kalman filtering by allowing different brain regions to have region-specific temporal noise properties. This improvement enables better separation of weak, deep sources from dominant cortical signals.

In both simulated EEG experiments and real EEG recordings from an epilepsy patient, RSTKF outperformed LORETA and the standard STKF in accuracy, consistency, and spatial specificity especially for subcortical sources that are traditionally very difficult to detect. Importantly, RSTKF maintained robust performance even with fewer electrodes or lower-quality EEG data.

2.5.1 The regional spatiotemporal Kalman filter (RSTKF)

The RSTKF is an advanced EEG source localization method designed to improve the detection of deep (subcortical) brain activity.

Key features:

- 3D brain modeling: The brain is represented as a voxel-based 3D grid, divided into anatomical regions (e.g., thalamus, hippocampus) using atlas-based segmentation.
- Region-specific dynamics: Each brain region is assigned unique dynamic noise parameters that reflect differences in neural signal characteristics across regions.
- Improved realism: Unlike STKF, which assumes uniform dynamics across the brain, RSTKF incorporates the brain's inherent modularity and heterogeneity, allowing better modeling of both cortical and deep sources.

This region-specific parameterization significantly enhances the ability to isolate and localize weak sources in deep structures such as the thalamus or hippocampus.

2.5.2 Development and validation of RSTKF

- **Development phase:**
  - Simulated EEG data were generated from single sources with known locations, orientations, and strengths.
  - Simulations used the realistically-shaped, three-compartment head model, based on the Boundary Element Method (BEM) for accurate forward modeling.
- **Validation phase:**
  - Applied RSTKF to single-trial interictal spikes from a presurgical epilepsy patient.
  - Performance metrics included localization accuracy, spatial focus, and consistency across spikes.

2.5.3 Achievements with RSTKF

- Introduced a new EEG source localization method with spatially heterogeneous model parameters, enabling accurate representation of multiple brain regions.
- Demonstrated superior accuracy and spatial resolution compared to LORETA and STKF, particularly for deep sources (e.g., putamen, thalamus).

- Successfully applied to clinical epilepsy data, achieving:
  - Consistent and accurate localization
  - Correct lateralization of all spikes at both onset and peak times

2.6 Robustness of Deep Learning-Based EEG Source Imaging Under Varying Electrode Configurations

Electroencephalography (EEG) is a non-invasive tool that monitors brain activity. However, finding the exact source of signals is difficult due to problems like volume conduction and limited electrode coverage. Electrophysiological source imaging (ESI) aims to solve this by estimating where brain signals come from. It usually works best with high-density electrode arrays, such as 64 or 75 channels. In settings with fewer resources, it's common to have lower electrode counts, 32 or fewer. Traditional methods like sLORETA and LCMV often result in larger errors in these situations. The DeepSIF method, which is based on deep learning, addresses this issue by learning complex brain patterns. This approach may provide solid performance with different electrode configurations, including 16, 21, 32, 64, and 75 channels [24].This flexibility could make ESI more accessible and practical.

2.6.1 Leadfield Matrix Approach

A critical component of DeepSIF is the leadfield matrix, which maps brain signals to electrode measurements. For different configurations (16, 21, 32, 64, and 75 channels) the leadfield matrices are derived by starting with a full 75-channel setup and downsampling to match the target electrode count. This consistent approach ensures accurate signal modeling across setups, akin to adapting a recipe to fewer ingredients while maintaining quality. It allows DeepSIF to be tested across a spectrum of electrode densities, from sparse 16-channel arrays to dense 75-channel setups.

2.6.2 Performance

Studies show DeepSIF is robust and reliable across these electrode configurations, outperforming sLORETA and LCMV, especially with fewer electrodes. In epilepsy patients, DeepSIF accurately identifies source locations, surpassing traditional methods even with as few as 16 channels. This makes it a promising tool for diverse clinical settings.

2.7 Research Gaps

2.7.1 Transformer-Based Architecture

Most methods continue to rely on traditional CNNs or RNNs that struggle with complex temporal patterns in multi-channel EEG data, despite facing their own limitations in handling data scarcity and noise common in EEG applications. We propose using Transformers because they are better suited for large datasets than other methods, potentially offering improved performance on large-scale datasets compared to other neural architectures. This gap hinders the potential for more accurate dynamic source tracking and multimodal integration, as Transformers could better model inter-channel correlations and spatial-temporal hierarchies without the limitations of recurrent processing, if challenges related to data size and computational demands are addressed [34].

2.7.2 Identify Multiple Source Activations in Source Imaging

Real-world brain activity often involves simultaneous activations from multiple neural sources, particularly in conditions like epilepsy or during multifaceted cognitive processes. Current EEG source imaging techniques predominantly focus on single or sparse source assumptions, leading to inaccuracies in disentangling overlapping signals and reduced ecological validity. This research gap limits the clinical applicability of models, as it overlooks the need for robust algorithms capable of identifying, separating, and localizing multiple concurrent sources while accounting for their interactions and noise interference.

2.7.3 Train with Less Data Size

Deep learning models for solving the EEG inverse problem typically demand enormous synthetic datasets, often exceeding terabytes in size, which imposes significant computational and storage burdens [8]. There is a notable gap in developing efficient training paradigms that achieve comparable or superior performance using substantially reduced data volumes, such as through advanced data augmentation, transfer learning, or knowledge distillation techniques. Addressing this would democratize access to high-quality models, especially for resource-limited research groups, and accelerate iteration in model development.

2.7.4 Analysis with Changing Source in Same Signal with Time

EEG signals are inherently dynamic, with neural sources potentially shifting locations, intensities, or orientations over time within a single recording segment. Existing approaches largely treat sources as static or rely on averaged snapshots, failing to capture these temporal evolutions and leading to incomplete reconstructions of brain activity [8]. This gap restricts applications in real-time monitoring or transient event analysis, such as seizure propagation, and calls for methods that incorporate time-varying source modeling to enhance temporal resolution and coherence in localization results.

2.7.5 CNN + Transformer-Based Architecture

Convolutional Neural Networks (CNNs) excel at extracting spatial features from EEG data, while Transformers are adept at handling sequential dependencies, yet hybrid architectures combining these strengths have not been extensively explored for EEG source localization. This represents a key research gap, as such integration could yield more robust spatio-temporal representations, outperforming standalone models in noisy or variable conditions. Without this, current frameworks miss opportunities to leverage complementary capabilities for improved accuracy, generalization, and efficiency in inverse problem solutions.

2.7.6 Finding General Method Without Depending on the Channels of Leadfield Matrix

Most EEG source localization methods are tightly coupled to specific leadfield matrices, which depend on electrode channel counts, head models, and sensor configurations, resulting in poor adaptability across different EEG systems or subjects [8]. There is a significant gap in creating generalizable approaches that operate independently of leadfield channel specifics, perhaps through adaptive normalization, meta-learning, or channel-agnostic feature encoding. This would enable broader deployment in diverse clinical settings, reducing the need for custom recalibrations and enhancing cross-dataset performance.

3 PROPOSED METHODOLOGY

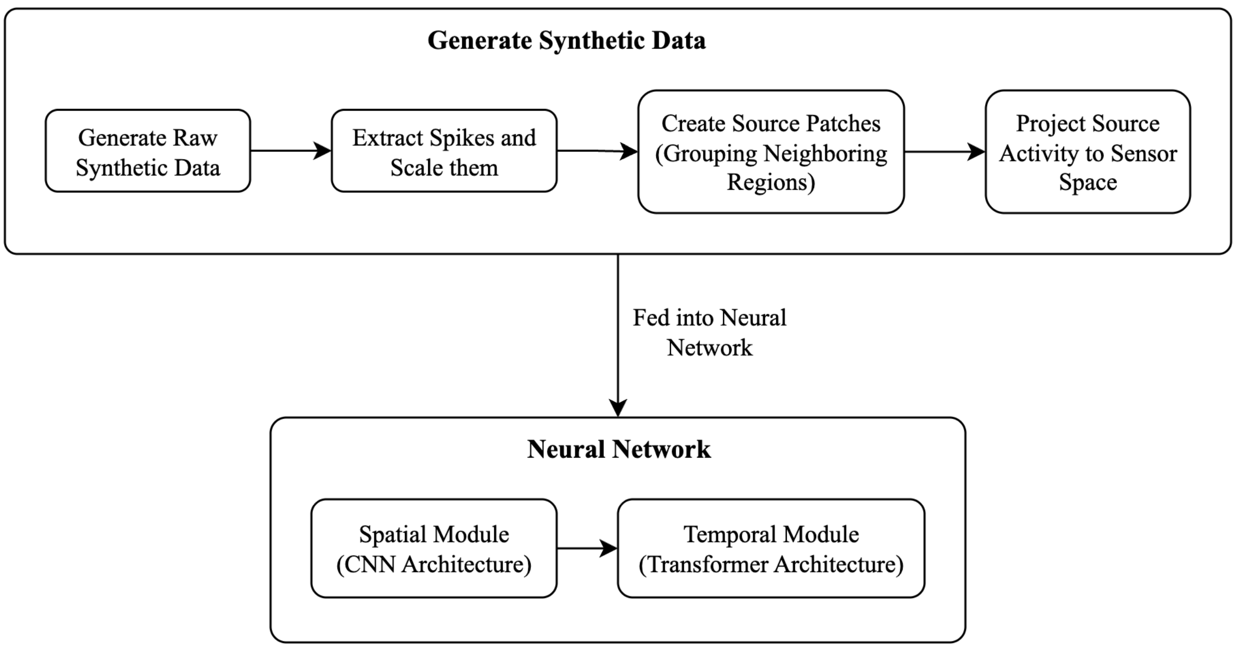

Fig. 4: Proposed method overview

We are proposing a new method for efficient and realistic synthetic data generation using Python libraries. After preprocessing, the data will be fed into our proposed model, which is developed using CNNs and Transformers. A high-level overview of our proposed method is shown in Fig. 4

3.1 Synthetic Data Generation

3.1.1 Raw Data Generation Using The Virtual Brain (TVB) (Step 1)

We simulate EEG data using The Virtual Brain (TVB)'s(for more details, see Appendix B). implementation of the Jansen-Rit model, which emulates the dynamics of cortical columns by capturing both excitatory and inhibitory neural interactions. The simulations are grounded in a biologically realistic structural connectivity matrix encompassing 998 brain regions, serving as the anatomical scaffold for activity generation. Each region is simulated for 50 seconds with a 1.0 ms time step, and the resulting activity is segmented into 10 files of 5 seconds each. This results in 30 files per region, amounting to approximately 20 GB of data in total. The outcome is a computationally efficient yet biologically rich dataset, suitable for downstream tasks such as spike detection and source localization.

This optimization reduces storage requirements and computational time by more than 10× while preserving the temporal resolution necessary for downstream spike extraction and source localization tasks.

Table 4: Comparison between DeepSIF Default and Optimized Simulation Parameters

Feature	**DeepSIF Default**	**Optimized Setup**
Simulation	200 s	50 s
Step Size	0.5 ms	1 ms
Files per Region	20	10
Total Dataset Size	∼4 TB	∼20 GB
Simulation Time (Total)	∼6000 hrs	∼500 hrs

As shown in Table 4 the optimization was adopted not only to reduce computational cost and storage, but also to enable the generation of longer-duration spikes, which are more suitable for analysis while still maintaining the essential signal characteristics with no significant loss in accuracy.

3.1.2 Spike Extraction (Step 2)

The goal here is to extract distinct spike-like neural events from each simulation that can be used to generate labeled input-output pairs for model training.

As shown in Figure 5 preprocessing pipeline includes:

- Concatenate simulation chunks.
- Remove the first 1000 time points to discard transient artifacts.
- Downsample by a factor of 4.

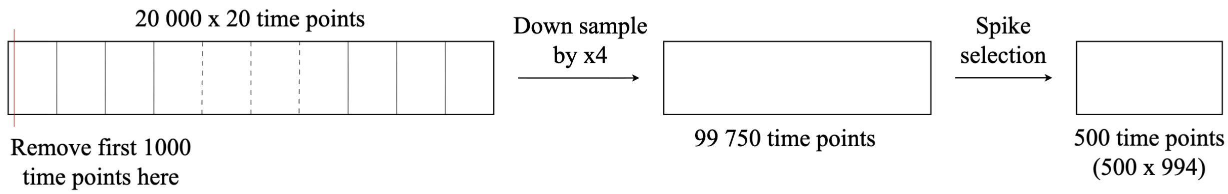

Fig. 5: Spike selection workflow

Spike Extraction Logic:

1. Thresholding: An amplitude threshold of 8 is applied; all values below are zeroed.

2. Local Maxima Selection: Peaks are sorted by time, ignoring unstable boundaries.

3. Region-Specific Filtering:

   - Rule 1: Only spikes originating from the target brain region for the current iteration are kept.
   - Rule 2: A 900-sample window around the spike must be free of significant activity from any other brain region, ensuring a single, clear source.

The downsampled signal used for spike extraction is shown in Figure 6, while the corresponding extracted spike is depicted in Figure 7.

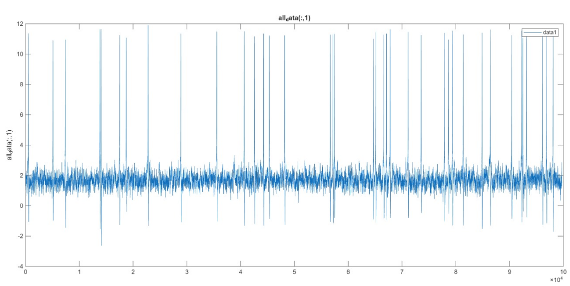

Fig. 6: Downsampled signal

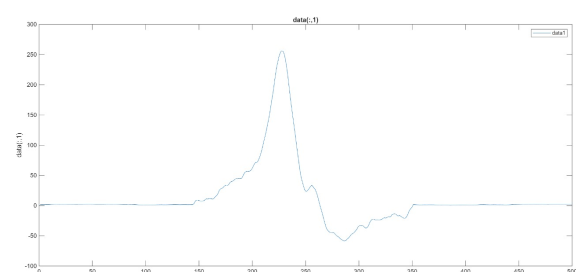

Fig. 7: Extracted spike

A 500-sample window is extracted around each spike and scaled to the desired Signal-to-Noise Ratio (SNR).

3.1.3 Synthetic Source Generation for Proposed Model Training (Step 3)

To train the proposed model, we must simulate where in the brain spikes occur. This requires creating realistic spatial source patches–groups of brain regions that act as a single coherent activation zone.

- Region Patch Construction

  - Based on the `fsaverage5` cortical atlas (994 regions), each synthetic source patch is centered on a selected region and expanded outward using a region-growing algorithm.
  - Patch growth is constrained by directionality, anatomical neighborhood, and smoothing heuristics, ensuring spatial coherence.

- Multi-Variant Sampling
  - In multi-source setups (e.g., `n_sources = 2`), additional patches are selected to simulate overlapping or separate brain activations.

- Signal Assignment & Control
  - Spike Assignment: Each patch is linked to a spike waveform extracted in Step 2.
  - SNR Scaling: Signals are scaled to predefined SNR levels (5–20 dB) using forward model projections.
  - Magnitude Decay: A Gaussian decay function models the drop in signal strength from the patch center to outer regions, mimicking realistic signal spread.

This process generates thousands of labeled examples that replicate real EEG signal behavior under controlled, interpretable conditions.

3.2 A Hybrid CNN-Transformer Architecture for Spatio-temporal Source Imaging

This research introduces a novel deep learning methodology for electrophysiological source imaging (ESI). Our approach is built upon a new hybrid architecture that marks a significant departure from previous data-driven frameworks. While prior models have successfully used recurrent architectures to model temporal brain dynamics, the sequential nature of this processing introduces specific limitations.

Our proposed architecture incorporates a Convolutional Neural Network (CNN) for spatial feature extraction. We posit that CNNs are fundamentally better suited for this task than the fully connected layers used in prior models. Their use of convolutional filters enforces weight sharing and builds in translational invariance, allowing the model to efficiently learn to recognize spatial patterns in the scalp topography regardless of their precise location. This inductive bias, absent in fully connected architectures, makes the model more parameter-efficient and robust against overfitting.

For temporal modeling, the architecture integrates a Transformer-based module, moving beyond the sequential processing limitations of recurrent models. The Transformer offers two distinct advantages. First, its core mechanism, multi-headed self-attention [10], allows the model to process all time points in a sequence simultaneously. This creates a global context and enables the direct modeling of relationships between any two points in the time series, offering a more powerful framework than the memory-gated approach of an

LSTM. Second, this parallel processing makes Transformers exceptionally well-suited for modern hardware, leading to greater computational efficiency during training compared to the inherently sequential nature of recurrent networks.

3.2.1 Architectural Design

As shown in Fig. 8, the proposed framework is a modular deep neural network designed to transform multi-channel scalp EEG data into estimates of underlying brain source activity.

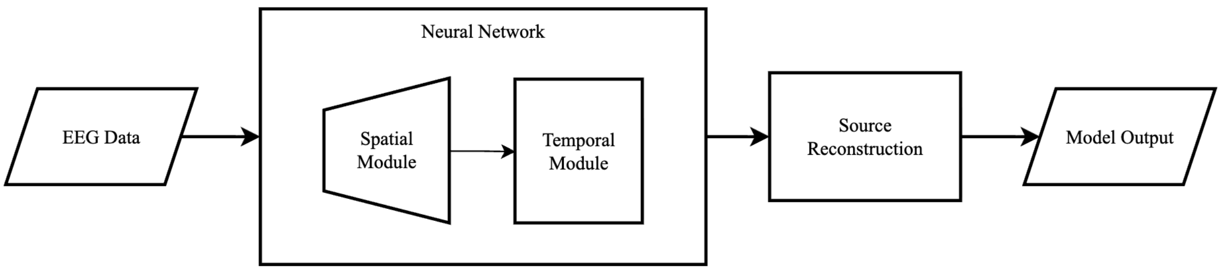

Fig. 8: High-Level diagram of proposed architecture

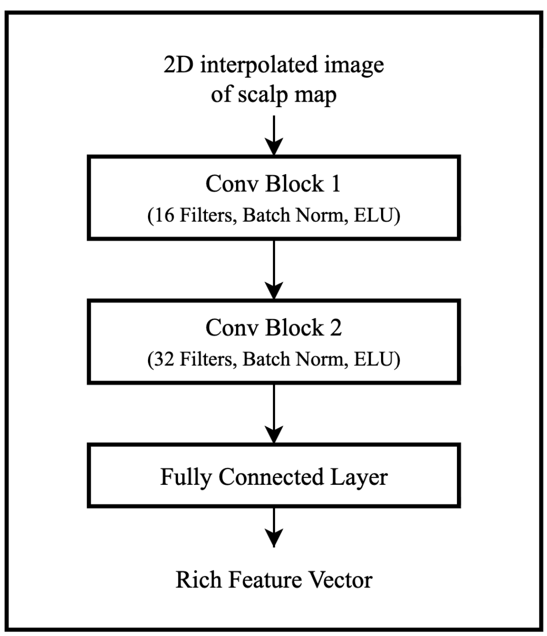

Fig. 9: Spatial module

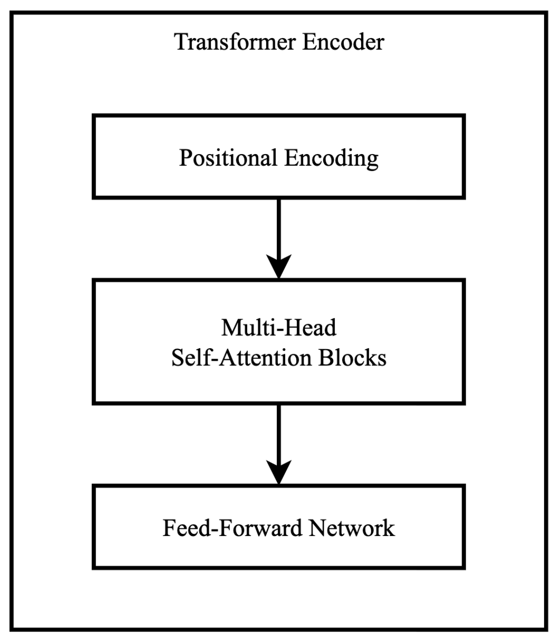

Fig. 10: Temporal module

Input Preprocessing (2D Topographical Map Generation):

To make the EEG data suitable for convolutional operations, the raw sensor measurements must be converted from a sparse set of locations into a spatially structured 2D grid. This process ensures that the spatial relationships between electrodes are preserved and represented in a format that a CNN can effectively process. This approach has been successfully validated in prior work, such as the ConvDip model [7], which interpolated EEG data into a low-resolution 7x11 grid to serve as its input. While the ConvDip paper confirms this step, it does not specify the exact interpolation algorithm used.

1. **Projection of Electrode Coordinates**
   The initial step involves projecting the 3D coordinates of the EEG electrodes onto a 2D plane. We will employ a standard cartographic method, such as an *Azimuthal Equidistant Projection*. This method projects the semi-spherical arrangement of scalp electrodes onto a 2D circular view, accurately preserving the relative distances and angular positions of each electrode from the center of the head.

2. **Investigation of Spatial Interpolation Methods**
   Once the electrodes are projected onto the 2D plane, a regular grid (e.g., 64×64 pixels) will be defined. Since the electrodes only occupy a few points on this grid, spatial interpolation is required to estimate the voltage values for the grid points in between. We will implement and evaluate the following standard methods:

   - Bilinear Interpolation: A computationally efficient method that estimates the value at a grid point by taking a weighted average of the four nearest known electrode locations. While fast, it can sometimes produce less smooth, more "blocky" topographies.
   - Bicubic Interpolation: A more complex method that considers a larger 4×4 neighborhood of 16 surrounding electrode points. It fits a cubic polynomial to these points, resulting in a smoother and more continuous surface compared to bilinear interpolation.
   - Radial Basis Function (RBF) Interpolation: A powerful method well-suited for scattered data like EEG electrodes. It models the potential map as a sum of basis functions, each centered on an electrode, where the function's influence decays with distance. This approach typically produces smooth and physically plausible topographical maps.

The output of this process for each time point is a 2D matrix, where each pixel's value corresponds to the interpolated electrical potential. By comparing the performance of our CNN with inputs generated by these different methods, we will determine the optimal preprocessing pipeline for our final model. This topographical map serves as the direct input to the Spatial Feature Extractor.

Spatial Feature Extractor:
For the crucial task of spatial feature extraction, our proposed architecture will employ a Convolutional Neural Network (CNN) module as shown in Fig. 9. This module will serve as the foundational stage of the network, engineered to process the 2D topographical scalp maps derived from multi-channel EEG data for each individual time step. The rationale for this design is that CNNs are fundamentally better suited than fully connected architectures for learning the complex spatial hierarchies in EEG data, which is the first critical step in

solving the inverse problem.

The proposed design of this module is specified as follows:

1. **Convolutional Blocks**
   The core of the module will be composed of several stacked convolutional blocks. Each block is engineered to extract increasingly complex spatial features. Our design for each block consists of a 2D Convolutional Layer, a Batch Normalization Layer, and an ELU Activation Function.

2. **Strategic Omission of Pooling Layers**
   A critical specification of our design is the deliberate omission of any pooling layers. For the task of source localization, where the goal is to determine the exact origin of a signal, preserving spatial fidelity is paramount.

3. **Feature Vectorization**
   Following the final convolutional block, a Flatten layer will be implemented. This layer's function is to transform the final 2D feature maps into a single, high-dimensional feature vector, which serves as a rich input to the subsequent temporal processing module.

Positional Encoding:
A critical component for a Transformer-based model is the Positional Encoding layer. Because Transformers process all data points in parallel, they lack an inherent sense of sequence order. To provide this crucial temporal information, we will add a pre-calculated positional vector to each feature vector as can be seen in Fig. 10. This is achieved using sinusoidal functions, which "timestamp" each data point and allow the model to utilize the relative positioning of events in the time series.

Temporal Module (Transformer Encoder):
This module is the core innovation of our methodology. The sequence of positionally-encoded feature vectors, which now contains both spatial information from the CNN and temporal context from the positional encoding, will be fed into a multi-layer Transformer Encoder as illustrated previously in Fig. 10. Each layer of the encoder performs two main operations:

- Multi-Head Self-Attention: This mechanism allows the model to weigh the importance of all other time steps when interpreting a single time step. It enables the framework to identify and model complex relationships between neural events, regardless of their distance from each other in the sequence.

- Position-wise Feed-Forward Network: Following the attention step, a standard feed-forward network is applied to each time step's representation for further non-linear processing, refining the learned features.

Crucially, each of these two sub-layers (attention and feed-forward) within an encoder layer also has a residual connection and is followed by layer normalization, which helps stabilize the learning process and enables the effective training of a deep stack of these layers.

Final Output Layer:
After passing through the Transformer Encoder, the resulting sequence of highly contextualized feature vectors is fed into a final linear (fully connected) layer. This layer projects the feature representation of each time step onto the desired output dimension, corresponding to the estimated activity in a predefined number of brain source regions.

3.2.2 Expected Advantages and Evaluation

The proposed hybrid architecture is strategically designed to overcome the key limitations of previous data-driven models by synergizing the distinct strengths of its specialized components. We anticipate the following advantages over prior works that relied on fully connected or recurrent architectures:

- Robust Spatial Feature Extraction: The CNN's inherent ability to learn topographical patterns is expected to provide more robust spatial features, making the model less sensitive to variations in noise and electrode placement.
- Superior Modeling of Temporal Dependencies: The self-attention mechanism is expected to provide a more robust method for capturing both short-range and long-range patterns in dynamic brain signals.
- Enhanced Performance: This improved modeling capability is hypothesized to lead to higher accuracy in localizing brain sources and estimating their time courses, resulting in improved validation metrics.
- Computational Efficiency: The parallel nature of the Transformer architecture is well-suited for modern hardware (GPUs/TPUs) and can lead to more efficient training compared to sequential models.

The performance of the proposed framework will be rigorously evaluated against conventional source imaging methods and other state-of-the-art deep learning architectures using a combination of synthetic and real-world human EEG datasets.

3.3 Model Training & Evaluation

The training process will utilize the synthetic EEG data generated by TVB library, combined with a limited amount of real data. The model will be pretrained on synthetic datasets to initialize weights, followed by fine-tuning on real EEG data to optimize performance. Training will be conducted using a supervised learning approach with a focus on minimizing localization errors.

Evaluation will involve testing the model on both synthetic and real EEG datasets, measuring accuracy and robustness using metrics such as localization error and correlation with ground truth. The model's performance will be assessed across different noise levels and electrode configurations to ensure generalizability. Fig. 11 illustrates the model evaluation process, where brain regions are segmented, matched with the predicted region indices, and compared against clinical findings.

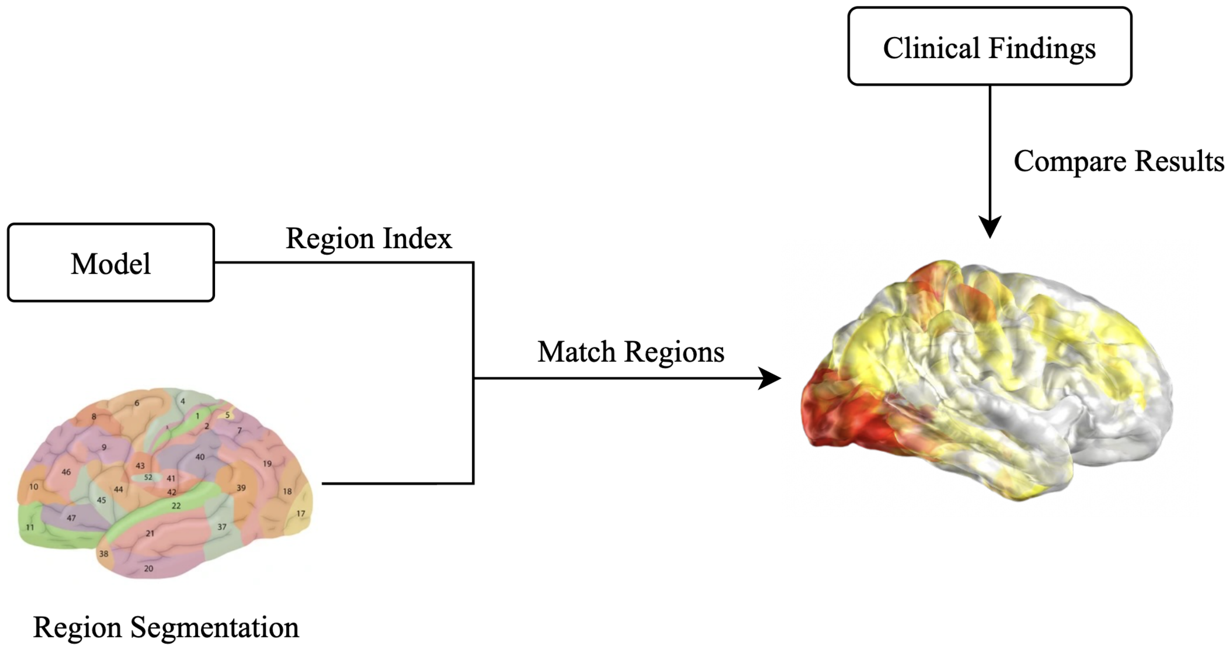

Fig. 11: Overview of the Model Evaluation Method

3.3.1 Datasets Hope to Use for Evaluate the Model

To validate the proposed framework, its performance will be tested on a diverse set of public datasets. We will use a combination of synthetic and real-world EEG data to assess the model's accuracy and generalizability across different conditions.

- The Open Database of Epileptic EEG with MRI real epilepsy patients dataset [35].
- MNE-Python toolbox healthy patients synthetic EEG data [36].
- LOCALIZE-MI: an open source dataset of simultaneous intracerebral stimulation and HD-EEG in humans [37].
- An open-access dataset of naturalistic viewing using simultaneous EEG-fMRI [38].

4 FEASIBILITY STUDY

This section evaluates the practicality of the proposed project by analyzing its key resource and technical requirements. The following sections assess the necessary physical infrastructure, time constraints, software tools, and computational power to confirm that the proposed work is viable and can be successfully completed.

4.1 Data Availability

The proposed method uses both public and synthetic EEG datasets. Public datasets, which include both epileptic and healthy subjects, are available for evaluation. This ensures reproducibility and allows for comparison with established standards. For model training, synthetic EEG data with known sources will be created. This enables controlled experiments and helps optimize deep learning models without the limitations of small annotated datasets.

4.2 Tools and Frameworks

The project will use a mix of established neuroimaging and deep learning tools. EEG preprocessing and source simulations will be done using MNE-Python, TVB, and MATLAB. These tools are popular and well-supported in the neuroimaging community. Development and training of the deep learning model will take place using PyTorch or TensorFlow. This choice offers flexibility for implementing the synthetic data generation method and CNN and Transformer architectures and other leading approaches.

4.3 Computational Resources

The computational requirements are moderate. Synthetic data generation and model training can be conducted on systems with standard GPUs and CPUs.We will use data management techniques like downsampling and chunking to make memory use more efficient and lower the computational load. This will help us experiment effectively without needing advanced computing clusters.

4.4 Research Support

The proposed study relies on established EEG source imaging methods. These methods offer reliable standards for evaluating models. Existing literature also shows that CNN-Transformer architectures are effective for EEG analysis. Furthermore, the TVB Python library can generate synthetic EEG data. This mix of methods and literature provides a solid base for the proposed research.

5 RESEARCH TIMELINE

The following Gantt chart in Fig. 12 outlines the projected timeline and key milestones for this research project. It provides a detailed breakdown of the tasks, their expected durations, and their sequential dependencies, ensuring a systematic and timely progression through all phases of the study. Please note that this schedule is a dynamic plan; it will gradually go through updates to reflect the evolving nature and needs of the research process.

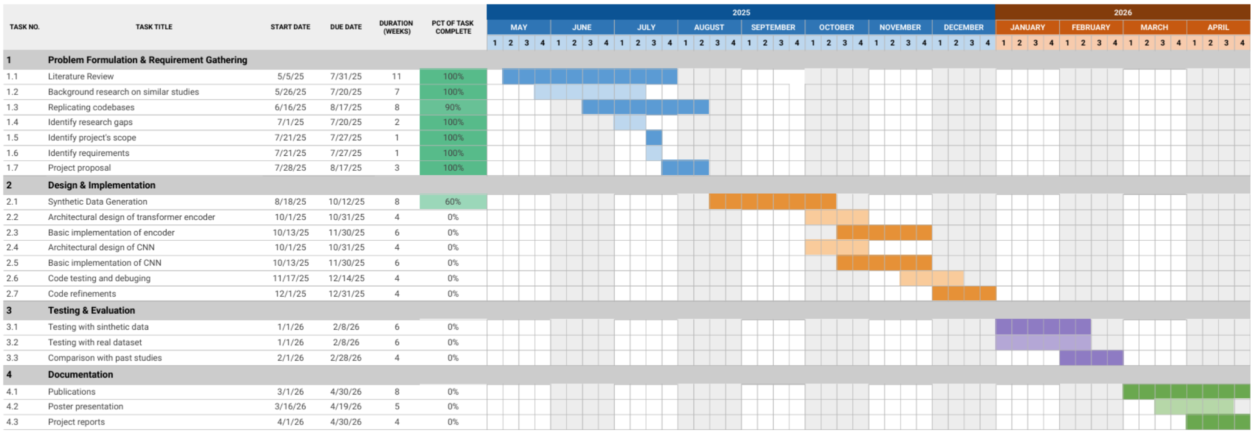

Fig. 12: Research Timeline

6 CONCLUSION

In summary, our research focuses on two main challenges in previous EEG source localization methods, the high resources needed for generating synthetic data and the limited accuracy of current deep learning models. Traditional and deep learning methods depend on large labeled datasets, which are often hard to come by in real-world situations. This makes synthetic data generation necessary. However, previous methods typically use computationally heavy generation processes. Our proposed method offers a more resource-efficient solution, allowing for the creation of enough training data using regular computing resources.

Additionally, we present a new deep learning model that combines a CNN-based spatial module with a Transformer-based temporal module. The CNN part uses convolutional kernels to capture spatial scalp patterns through weight sharing and translational invariance, providing better parameter efficiency and robustness than fully connected layers. The Transformer module is used to model temporal dependencies, effectively handling sequential information, especially with larger datasets. Together, these contributions create a more efficient and accurate framework for EEG source localization.

Ultimately, by improving the reliability and strength of the EEG inverse solution across different electrode setups and multiple source activations, our approach has great potential to aid clinical diagnosis. More dependable source imaging can lead to earlier and less invasive detection of neurological issues like epilepsy, which can improve treatment planning and increase practical uses of EEG in healthcare.

References

[1] H. Hallez, B. Vanrumste, R. Grech, J. Muscat, W. D. Clercq, A. Vergult, Y. D'Asseler, K. P. Camilleri, S. G. Fabri, S. V. Huffel, and I. Lemahieu, "Review on solving the forward problem in eeg source analysis," *Journal of NeuroEngineering and Rehabilitation*, vol. 4, no. 1, p. 46, Dec. 2007, published: 20 Dec. 2007. [Online]. Available: https://doi.org/10.1186/1743-0003-4-46

[2] C. A. Brebbia and J. Domingues, *Boundary Elements: An Introductory Course*. Computational Mechanics Publications, 1992. [Online]. Available: https://www.boundaryelements.com/index.php/resources-section/open-access-material/26-boundary-elements-an-introductory-course

[3] W. K. Liu, S. Li, and H. S. Park, "Eighty years of the finite element method: Birth, evolution, and future," *Archives of Computational Methods in Engineering*, vol. 29, no. 4, pp. 4431–4453, 2022. [Online]. Available: https://doi.org/10.1007/s11831-022-09740-9

[4] J. Peiró and S. Sherwin, "Finite difference, finite element and finite volume methods for partial differential equations," *Computational Mechanics*, vol. 30, no. 3, pp. 235–246, 2002. [Online]. Available: https://doi.org/10.1007/s004660200283

[5] R. Grech *et al.*, "Review on solving the inverse problem in eeg source analysis," *Journal of NeuroEngineering and Rehabilitation*, vol. 5, no. 1, p. 25, Dec. 2008. [Online]. Available: https://jneuroengrehab.biomedcentral.com/articles/10.1186/1743-0003-5-25

[6] S. Allouch, A. Kabbara, J. Duprez, M. Khalil, J. Modolo, and M. Hassan, "Effect of channel density, inverse solutions and connectivity measures on eeg resting-state networks reconstruction: A simulation study," *NeuroImage*, vol. 271, p. 120006, May 2023. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811923001520

[7] L. Hecker, R. Rupprecht, L. T. V. Elst, and J. Kornmeier, "Convdip: A convolutional neural network for better EEG source imaging," *Frontiers in Neuroscience*, vol. 15, p. 569918, Jun. 2021, published: 09 Jun. 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2021.569918/full

[8] R. Sun, A. Sohrabpour, G. A. Worrell, and B. He, "Deep neural networks constrained by neural mass models improve electrophysiological source imaging of spatiotemporal brain dynamics," *Proceedings of the National Academy of Sciences*, vol. 119, no. 31, p. e2201128119, 2022. [Online]. Available: https://doi.org/10.1073/pnas.2201128119

[9] X. Liang, Y. Liu, Y. Yu, K. Liu, Y. Liu, and Z. Zhou, "Convolutional neural networks with a topographic representation module for eeg-based brain-computer interfaces," *arXiv*, Aug. 2022, arXiv:2208.10708. [Online]. Available: https://arxiv.org/abs/2208.10708

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017. [Online]. Available: https://arxiv.org/abs/1706.03762

[11] B. Wang, X. Fu, Y. Lan, L. Zhang, W. Zheng, and Y. Xiang, "Large transformers are better eeg learners," *arXiv*, Apr. 2024, arXiv:2308.11654. [Online]. Available: https://arxiv.org/abs/2308.11654

[12] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18, no. 6, pp. 14–30, Nov. 2001, published: Nov. 2001. [Online]. Available: https://doi.org/10.1109/79.962275

[13] A. Hashemi, M. B. Shamsi, M. A. Bayati, H. Soltanian-Zadeh, and S. Sanei, "Bayesian uncertainty quantification in eeg source imaging," *NeuroImage*, vol. 237, p. 118158, 2021, published: 1 Aug. 2021. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2021.118158

[14] W. Weng, Y. Gu, S. Guo, Y. Ma, Z. Yang, Y. Liu, and Y. Chen, "Self-supervised learning for electroencephalogram: A systematic survey," *arXiv*, vol. 2401.05446, 2024, arXiv preprint. [Online]. Available: https://doi.org/10.48550/arXiv.2401.05446

[15] F. P. Carrle, Y. Hollenbenders, and A. Reichenbach, "Generation of synthetic eeg data for training algorithms supporting the diagnosis of major depressive disorder," *Frontiers in Neuroscience*, vol. 17, p. 1219133, 2023, published: 2 Oct. 2023, PMID: 37849893, PMCID: PMC10577178. [Online]. Available: https://doi.org/10.3389/fnins.2023.1219133

[16] L. Xu, M. Xu, Y. Ke, X. An, S. Liu, and D. Ming, "Cross-dataset variability problem in eeg decoding with deep learning," *Frontiers in Human Neuroscience*, vol. 14, p. 103, Apr. 2020, published online: 21 Apr. 2020, PMID: 32372929, PMCID: PMC7188358. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00103/full

[17] D. Bethge, P. Hallgarten, O. Özdenizci, R. Mikut, A. Schmidt, and T. Grosse-Puppendahl, "Exploiting multiple eeg data domains with adversarial learning," *arXiv*, vol. 2204.07777, 2022, arXiv preprint. [Online]. Available: https://doi.org/10.48550/arXiv.2204.07777

[18] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017. [Online]. Available: https://doi.org/10.1002/hbm.23730

[19] H. He, W. Wu, L. Li, Y. Zhang, and Y. Huang, "Deep learning for eeg inverse problem: A review," *Frontiers in Neuroscience*, vol. 16, p. 853525, 2022. [Online]. Available: https://doi.org/10.3389/fnins.2022.853525

[20] C. M. Michel, D. Brandeis, T. W. Koenig, L. R. Gianotti, and D. Lehmann, "Eeg source imaging," *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2195–2222, 2004. [Online]. Available: https://doi.org/10.1016/j.clinph.2004.06.001

[21] R. D. Pascual-Marqui, "Standardized low-resolution brain electromagnetic tomography (sloreta): technical details," *Methods and Findings in Experimental and Clinical Pharmacology*, vol. 24, no. Suppl D, pp. 5–12, 2002, pMID: 12575463. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/12575463/

[22] A. M. Dale, M. H. Fischl, and R. B. Sereno, "Dynamic statistical parametric mapping: Combining fmri and meg for high-resolution imaging of cortical activity," *Neuron*, vol. 26, no. 1, pp. 55–67, 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0896627300811381

[23] C. M. Michel and D. Brunet, "EEG source imaging: A practical review of the analysis steps," *Frontiers in Neurology*, vol. 10, p. 325, Apr. 2019, published: 04 Apr. 2019. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fneur.2019.00325/full

[24] J. Rong, R. Sun, B. Joseph, G. Worrell, and B. He, "Deep learning-based eeg source imaging is robust under varying electrode configurations," *Clinical Neurophysiology*, vol. 175, p. 10730, 2025. [Online]. Available: https://doi.org/10.1016/j.clinph.2025.04.009

[25] R. Sun, W. Zhang, A. Bagić, and B. He, "Deep learning based source imaging provides strong sublobar localization of epileptogenic zone from meg interictal spikes," *NeuroImage*, vol. 281, p. 120366, Sep. 2023, published: 15 Sep. 2023. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2023.120366

[26] R. Sun, "Deep learning-based source imaging improves spatiotemporal imaging of epileptic sources," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA, 2023. [Online]. Available: https://www.proquest.com/openview/0041E_10985.pdf

[27] M. Morik, A. Hashemi, K.-R. Müller, S. Haufe, and S. Nakajima, "Enhancing brain source reconstruction through physics-informed 3d neural networks," *arXiv preprint*, vol. arXiv:2411.00143, 2024. [Online]. Available: https://arxiv.org/abs/2411.00143

[28] W. Cui, W. Jeong, P. Thölke, T. Medani, K. Jerbi, A. A. Joshi, and R. M. Leahy, "Neuro-gpt: Towards a foundation model for eeg," *arXiv preprint*, vol. arXiv:2311.03764, 2023. [Online]. Available: https://arxiv.org/abs/2311.03764

[29] W.-B. Jiang, L.-M. Zhao, and B.-L. Lu, "Large brain model for learning generic representations with tremendous eeg data in bci (labram)," *arXiv preprint*, vol. arXiv:2405.18765, 2024. [Online]. Available: https://arxiv.org/abs/2405.18765

[30] S. Wang, T. Zhou, Y. Shen, Y. Li, G. Huang, and Y. Hu, "Generative ai enables EEG super-resolution via spatio-temporal adaptive diffusion learning," *arXiv preprint*, vol. arXiv:2407.03089, 2024. [Online]. Available: https://arxiv.org/abs/2407.03089

[31] M. Khosravi, A. Kalhor, F. Bahrami, and B. Moshiri, "Solving the inverse problem for eeg signals when learning a new motor task using gru neural network," in *2023 31st International Conference on Electrical Engineering (ICEE)*, Tehran, Iran, 2023, pp. 910–914. [Online]. Available: https://ieeexplore.ieee.org/document/10334704

[32] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014, eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8). [Online]. Available: https://arxiv.org/abs/1409.1259

[33] A. Pascarella, Y. Zhang, D. Marinazzo, J. Meuleman, J. Hajnal, L. Zollei, S. V. Huffel, M. Verhoye, S. Sunaert, W. V. Paesschen, B. Boon, H. Sahli, C. Philips, and M. D. Vos, "Source imaging of deep-brain activity using the regional spatiotemporal kalman filter," *NeuroImage*, vol. 268, p. 119856, Jan. 2023, published: 15 Jan. 2023. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2022.119856

[34] Transformer vs. LSTM: 4 key differences and how to choose. [Online]. Available: https://www.kolena.com/guides/transformer-vs-lstm-4-key-differences-and-how-to-choose/

[35] P. Zwoliński, M. Roszkowski, J. Zygierewicz, S. Haufe, G. Nolte, and P. J. Durka, "Open database of epileptic eeg with mri and postoperational assessment of foci – a real world verification for the eeg inverse solutions," *Neuroinformatics*, vol. 8, no. 4, pp. 285–299, Dec. 2010, published: Dec. 2010, PMID: 20872095, PMCID: PMC2974908. [Online]. Available: https://doi.org/10.1007/s12021-010-9086-6

[36] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. Hämäläinen, "Meg and eeg data analysis with mne-python," *Frontiers in Neuroscience*, vol. 7, pp. 1–13, 2013, published: 26 Dec. 2013, PMID: 24431986, PMCID: PMC3872725. [Online]. Available: https://doi.org/10.3389/fnins.2013.00267

[37] E. Mikulan, S. Russo, S. Parmigiani, S. Sarasso, F. M. Zauli, A. Cattani, M. Massimini, and A. Pigorini, "Localize-mi: Simultaneous human intracerebral stimulation and hd-eeg — ground-truth for source localization methods," Human Brain Project Neuroinformatics Platform, 2020, dataset; 61 sessions in 7 human subjects; includes EEG, MRI, electrode positions, forward models etc. [Online]. Available: https://doi.org/10.25493/NXN2-05W

[38] Q. K. Telesford, E. Gonzalez-Moreira, T. Xu, Y. Tian, S. J. Colcombe, J. Cloud, B. E. Russ, A. Falchier, M. Nentwich, J. Madsen, L. C. Parra, C. E. Schroeder, M. P. Milham, and A. R. Franco, "An open-access dataset of naturalistic viewing using simultaneous eeg-fmri," *Scientific Data*, vol. 10, no. 1, p. 554, 2023, published: 23 Aug. 2023; PMID: 37612297; PMCID: PMC10447527. [Online]. Available: https://doi.org/10.1038/s41597-023-02458-8

[39] M. Schirner, L. Domide, D. Perdikis, P. Triebkorn, L. Stefanovski, R. Pai, P. Prodan, B. Valean, J. Palmer, C. Langford, A. Blickensdörfer, M. van der Vlag, S. Diaz-Pier,

A. Peyser, W. Klijn, D. Pleiter, A. Nahm, O. Schmid, M. Woodman, L. Zehl, J. Fousek, S. Petkoski, L. Kusch, M. Hashemi, D. Marinazzo, J.-F. Mangin, A. Flöel, S. Akintoye, B. C. Stahl, M. Cepic, E. Johnson, G. Deco, A. R. McIntosh, C. C. Hilgetag, M. Morgan, B. Schuller, A. Upton, C. McMurtrie, T. Dickscheid, J. G. Bjaalie, K. Amunts, J. Mersmann, V. Jirsa, and P. Ritter, "Brain simulation as a cloud service: The virtual brain on ebrains," *NeuroImage*, vol. 251, p. 118973, May 2022, published: 1 May 2022, Epub Feb 4 2022; PMID: 35131433. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2022.118973

[40] A. Rayi and N. I. Murr, "Electroencephalogram," In: StatPearls [Internet], 2025, updated: 3 Oct 2022, Ongoing publication. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK563295/

[41] S. J. Luck, *An Introduction to the Event-Related Potential Technique*. MIT Press, 2014, iSBN: 978-0262525855. [Online]. Available: https://mitpress.mit.edu/9780262525855/an-introduction-to-the-event-related-potential-technique/

[42] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012. [Online]. Available: https://doi.org/10.3390/s120201211

[43] M. Carpi and C. Liguori, "Sleep eeg in chronic insomnia disorder," *Clinical Neurophysiology*, vol. 176, p. 2110774, Aug. 2025, epub: 28 May 2025, PMID: 40460740. [Online]. Available: https://doi.org/10.1016/j.clinph.2025.2110774

[44] A. Chail, R. K. Saini, P. S. Bhat, K. Srivastava, and V. Chauhan, "Transcranial magnetic stimulation: A review of its evolution and current applications," *Indian Psychiatry Journal*, vol. 27, no. 2, pp. 172–180, 2018, published: Jul–Dec 2018, PMID: 31359968, PMCID: PMC6592198. [Online]. Available: https://doi.org/10.4103/ipj.ipj_88_18

Appendix A: Overview of EEG

A.1 What is EEG?

EEG is a non-invasive method that records the brain's electrical activity with electrodes placed on the scalp. These electrodes pick up voltage changes caused by postsynaptic potentials in cortical neurons. EEG offers high temporal resolution, typically in milliseconds, which makes it particularly useful for studying fast-changing neural processes. [40]

A.2 Physiological Basis

EEG signals mainly show the combined excitatory and inhibitory postsynaptic potentials of pyramidal neurons in the cerebral cortex. [41]
The brain's rhythmic activity is divided into frequency bands:

- Delta (0.5-4 Hz): Deep sleep, unconscious states.
- Theta (4-8 Hz): Drowsiness, meditation, memory processing.
- Alpha (8-13 Hz): Relaxed wakefulness, eyes closed.
- Beta (13-30 Hz): Active thinking, concentration.
- Gamma (>30 Hz): Cognitive and sensory processing.

A.3 Applications of EEG

EEG is widely used in both clinical and research domains:

- Clinical applications: epilepsy diagnosis, sleep disorders, monitoring brain death, intraoperative monitoring. [43]
- Research applications: cognitive neuroscience, brain-computer interfaces (BCIs), psychiatric disorder studies, source localization. [42]

A.4 Advantages and Limitations

- Advantages: High temporal resolution, non-invasive, inexpensive, widely available. [41]
- Limitations: Poor spatial resolution, sensitive to artifacts, requires advanced inverse modeling. [5]

A.5 Relevance to the EEG Inverse Problem

The EEG inverse problem involves figuring out the brain's neural sources based on EEG signals recorded from the scalp. Because this problem is tricky to solve, it requires complex math and computer methods. [23] This is key to improving the spatial resolution of EEG and increasing its use in clinical diagnosis and neuroscience research.

Appendix B: Overview of The Virtual Brain (TVB)

B.1 What is TVB?

The Virtual Brain (TVB) [39] is an open-source platform for neuroinformatics that focuses on large-scale brain network simulations. It combines structural and functional neuroimaging data, like MRI, DTI, fMRI, and EEG/MEG, to build personalized models of brain dynamics. TVB enables researchers to simulate neural activity on various spatial and temporal levels. This helps connect neural physiology with the brain signals we observe.

B.2 Core Features

- Structural Connectivity Integration: Uses individual or template-based connectomes to model anatomical connections between brain regions.
- Dynamic Neural Mass Models: Supports different neural population models, such as Jansen-Rit and Wong-Wang, to simulate cortical and subcortical activity.
- Forward Modeling: Generates simulated neuroimaging data, including EEG, MEG, and BOLD signals, from neural activity. This enables a direct comparison with empirical measurements.
- Parameter Exploration and Optimization: Helps tune model parameters to match empirical data. This allows for personalized brain modeling.
- High-Performance Computing Support: Can run simulations on local machines or distributed computing clusters for large-scale networks.

B.3 Applications

TVB has become a powerful tool in both research and clinical neuroscience.

- Understanding Brain Dynamics: Studying large-scale network interactions that support cognition, attention, and consciousness.
- Disease Modeling: Simulating conditions like epilepsy, stroke, and neurodegenerative disorders to understand their underlying mechanisms.
- Brain Stimulation Planning: Helping in planning transcranial magnetic stimulation (TMS) [44] or deep brain stimulation treatments.
- EEG/MEG Source Localization: Offering forward models to test and validate inverse problem solutions, improving source localization accuracy.

B.4 Advantages and Limitations

B.4.1 Advantages

- Integrates multimodal neuroimaging data for personalized models.
- Flexible neural mass and connectivity models.
- Supports both forward and inverse modeling.

B.4.2 Limitations

- Requires high-quality structural connectivity data for accurate simulations.
- Computationally intensive for large networks or long-duration simulations.
- Model parameters can be tricky to adjust and understand.

B.5 Relevance to EEG Inverse Problem

TVB is important for EEG source localization research because it can simulate realistic EEG signals from known neural sources. By providing synthetic but physiologically plausible data, TVB allows for the testing and training of deep learning models for EEG inverse solutions. This helps to improve spatial resolution and reliability across different electrode setups.